Technical Report 1062

Analysis of Battlefield Operating System (BOS) Statements for Developing Performance Measurement

Angelo Mirabella
U.S. Army Research Institute

April 1997

United States Army Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Director

Technical review by

George W. Lawton
Douglas Macpherson

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to : U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-STP, 5001 Eisenhower Ave., Alexandria, Virginia 22333-5600

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED (from. . . to)
1997, April	Final	March 1993-September 1996

4. TITLE AND SUBTITLE	5a. CONTRACT OR GRANT NUMBER
Analysis of Battlefield Operating System (BOS) Statements for Developing Performance Measurement	5b. PROGRAM ELEMENT NUMBER 0603007A
6. AUTHOR(S) Angelo Mirabella	5c. PROJECT NUMBER A793
	5d. TASK NUMBER 2122
	5e. WORK UNIT NUMBER H01
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: PERI-II 5001 Eisenhower Avenue Alexandria, VA 22333-5600	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Technical Report 1062

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT *(Maximum 200 words)*:

This study was an initial effort in a larger program to develop training feedback measures and decision support methodology (DSM) for selecting brigade training strategies. Essential to either purpose is a set of reliable and valid unit performance measures. Such measures have traditionally been derived from front-end analysis. But archival data from the National Training Center (NTC) make possible a new, complementary approach of deriving measures from exercise data. In this study Battlefield Operating System (BOS) Impact Statements were used to derive unit performance measurement concepts. It was found that Impact Statements can be used reliably to judge relative unit performance across exercises, can be related to mission outcome, and can be used as a basis for deriving improved measures for training feedback and training effectiveness.

15. SUBJECT TERMS

Battlefield Operating System (BOS) Impact statements Decision support methodology (DSM)
Training feedback measures National Training Center (NTC)

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	41	Dr. Angelo Mirabella (703) 617-0325

Technical Report 1062

Analysis of Battlefield Operating System (BOS) Statements for Developing Performance Measurement

Angelo Mirabella
U.S. Army Research Institute

Advanced Training Methods Research Unit
Robert J. Seidel, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

April 1997

Army Project Number
2O363007A793

Training Systems and Education

FOREWORD

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) conducts research on how to design unit training strategies and methods. Strategy refers to allocation and scheduling of resources across training events such as FTXs, SIMNET exercises, and combat training center (CTC) rotations. Strategies per se are developed by Army Training and Doctrine (TRADOC) proponent schools and by unit commanders. This program seeks ways to assist the schools and commanders in their efforts. Methods include techniques for providing training feedback. Training strategies and methods need reliable and valid unit measurement concepts and instruments for successful implementation.

Such concepts and instruments traditionally have been developed as part of front end analysis. The effort summarized by this report explores a new approach based on data from the National Training Center (NTC). It explores the potential value of using Battlefield Operating System Impact Statements from the NTC to derive performance measures.

ZITA M. SIMUTIS
Technical Director

EDGAR M. JOHNSON
Director

ACKNOWLEDGMENTS

The author wishes to thank the subject-matter experts who made this research possible: Tom Mastaglio (LTC,Ret), Edward A. Rock (MAJ,Ret), Jim Love (LTC,Ret), MAJ Peter Hayes, and LTC Robert Earhart.

ANALYSIS OF BATTLEFIELD OPERATING SYSTEM (BOS) STATEMENTS FOR DEVELOPING PERFORMANCE MEASUREMENT

EXECUTIVE SUMMARY

Research Requirement:

Define unit performance measurement concepts and formats from archival data at the National Training Center (NTC) for use in developing methods to (1) select training strategies, and (2) provide remedial feedback to trainees.

Procedure:

1. Select a sample of Battlefield Operating System Impact Statements for Rank-Order Analysis: Ten statements each for Intelligence, Command and Control (C2), and Maneuver BOSs were selected. 2. Analyze BOS Impact Statements for Elements of Analysis: A taxonomy of elements was produced for sub BOSs within each of the three BOSs, e.g., planning, preparation, execution. 3. Rank Order the Impact Statements for Relative Performance for each BOS and sub BOS: Six analysts used paired comparisons to rank order performance on the BOSs. 4. Compute correlations of rankings with a battle outcome measure - the METT-T score: Correlations were computed for each analyst, BOS, and sub BOS.

Findings:

BOS impact statements were used reliably to evaluate the relative effectiveness of unit performance across task force exercises and phases of battle. Overall rankings of the impact statements for Maneuver and C2 were shown to correlate with METT-T scores. Relative performance across exercises for the Preparation and Execution Phases of battle for Maneuver and C2 is related to mission outcome (METT-T score). Results for the Intelligence BOS were inconclusive. A unit assessment instrument can be derived from the BOS Impact Statement by reformatting it to reflect phases of battle and elements of analysis within those phases. In addition, a five-point measurement scale was suggested to support the preparation of training feedback.

Utilization of Findings:

The results provided a basic research foundation for a follow-up program to apply path analysis to the study of performance within and among BOSs. It also provides concrete recommendations for how to immediately improve the analysis of BOS performance at NTC.

ANALYSIS OF BATTLEFIELD OPERATING SYSTEM (BOS) STATEMENTS FOR DEVELOPING PERFORMANCE MEASUREMENT

CONTENTS

CONTENTS (Continued)

Page

LIST OF TABLES

LIST OF FIGURES

Analysis of Battlefield Operating System (BOS) Statements for Developing Performance Measurement

1.0 INTRODUCTION

1.1 Background

a. This report summarizes an initial effort in a larger program to develop decision support methodology (DSM) for selecting brigade training strategies, and methodology for providing feedback to trainees. Essential to either purpose is a set of reliable and valid unit performance measures. Such measures have traditionally been derived from front-end analysis. But archival data from the National Training Center makes possible a new, complementary approach of deriving measures from training exercise data.

b. In this study, Battlefield Operating System (BOS) Impact Statements were used to derive unit performance measurement concepts and training assessment formats. The U.S. General Accounting Office (1986) underscored the value of the National Training Center for such development. The concepts and formats could help improve measurement to support training strategy, training feedback, and the human factors components of combat models.

c. Data from the CTCs are potentially invaluable for developing and validating unit performance measures. After surveying the NTC data base it was decided to concentrate initially on the Battlefield Operating System (BOS) Impact Statements. Reasons for doing so are described below. For details on the BOS itself see TRADOC Pam 11-9 (Department of the Army, 1990).

d. The ARI program has focused on brigade (Bde) training and more recently on the Army's role in Joint Service Training. These thrusts imply a need for "rolled-up" performance measures. These could be compilations or derivations from the CTC data streams. The BOS statement is potentially useful for developing such measures. It summarizes strengths and weaknesses of exercises at NTC for seven major functions of battle. These include Intelligence, Maneuver, Command and Control, Fire Support, Air Defense, Mobility/Survivability, and Combat Service Support.

e. The impact statements "roll-up" performance problems and successes across echelons, phases of battle, and tasks, within each functional area. The Statements are summaries of training needs for use in After Action Reviews. They are intended to focus on specific performance problems of the unit undergoing an NTC rotation.

f. Nonetheless, they may provide a basis for identifying critical performance dimensions, developing indicators, and engineering a measurement subsystem. These three developments would support training feedback and training strategy DSM. The first step was to examine the measurement properties of the BOS statements. The next step was to derive measurement concepts and instruments.

1.2 Research Issues

a. Can BOS Impact Statements be used consistently to rank order unit performance across exercises?

b. Can BOS performance be related to mission outcome?

c. Can BOS statements be used to derive performance indicators?[1]

---

[1] These questions do not imply an assessment of the work of O/Cs or a critique of the way BOS impact statements are prepared by the O/Cs. The impact statements are designed to provide specific feedback to specific units as part of the take-home package. They do that very well. Properly, they are not designed for testing or for research and development purposes.

However, this project is an exploratory development effort to identify sources of CTC archival information which may provide a basis for generating measures that will be useful in doing BDE-level training strategy trade-offs. Those measures, when developed, would be designed for use mainly in simulated networking and distributed interactive simulations. They would provide a basis for linking 'home station' training to CTC rotations.

2.0 METHODOLOGY

2.1 Select Data Set for Rank-Order Analysis

a. The data set for the analysis included BOS impact statements from 10 defend-in-sector exercises and METT-T scores from those exercises. The METT-T score is an experimental outcome measure developed by the ARI Presidio of Monterey Field Unit. It is derived from blue and red force casualties, and number of weapon systems in control of an objective at the end of an exercise (Kierins, Atwood, and Root, 1990).[2]

b. Three categories of BOS were selected for preliminary study. These were Intelligence, Command and Control (C2), and Maneuver. They seemed most suitable for examining large-unit measurement issues and likely to correlate with the METT-T score.

2.2 Review BOS Impact Statements for Elements of Analysis

a. With the help of military subject matter experts, the impact statements were content analyzed to identify elements of analysis. These were inputs to, processes, and products out of each phase of battle. These elements (Appendices A, B, and C) were included in instructions to help raters assess the impact statements. The elements were clustered by major phase of battle within each BOS, e.g., planning and execution phases of command and control. Taxonomies of sub BOS statements and some of their key elements are presented in Appendices A, B, and C.

(1) Appendix A shows BOS sub functions derived from the Intelligence BOS impact statements with battle phases. These focus on intelligence planning by the S2 for recon, counter recon, counter recon battle, and main battle. In the execution phase they focus on S2 tracking and battle analysis. Here they also focus on data gathering and some combat engagements by land and air scouts.

---

[2] METT-T Score (Blue defending) = [% Blue surviving + (100 - % OPFOR surviving) + (100 - % OPFOR in Blue territory)]/3

METT-T Score (Blue attacking) = [% Blue surviving + (100 - % OPFOR surviving) + % Blue in OPFOR territory]/3

Example (Blue defending): at the end of exercise
20% Blue weapons (tanks, TOWs, APCs) remain.
80% OPFOR weapons remain.
60% OPFOR weapons have crossed Blue's defensive line.

METT-T = [20 + (100-80) + (100-60)]/3 = 27.

(2) Appendix B shows BOS subfunctions derived from the C2 BOS impact statements. These are embedded within Planning, Preparation, and Execution phases of battle. Planning comments center on the TF Operations Order (OpOrd) and the staff operations leading to it. Preparation comments deal with rehearsals and defensive preparations. Major execution problems are rearward passage of security forces and control of the counter reconnaissance and main battles.

(3) Appendix C shows BOS subfunctions derived from the Maneuver BOS impact statements. These are organized under Planning, Preparation, and Execution-Counter Recon and Execution-Main Battle. Planning for maneuver focuses on battle formations, synchronization of combat elements, and repositioning of combat elements for various contingencies. Preparation deals with rehearsing formations and movements, and preparing defensive positions. Execution is divided into the counter reconnaissance and main battles.

2.3 Rank Order the Impact Statements For Relative Performance

Six raters independently rank-ordered the 10 impact statements for relative performance on the three BOSs. They used the method of pair-wise comparisons (Appendix D). After ranking overall BOS performance, the SMEs further rank-ordered the impact statements by phase of battle: planning, preparation, and execution. Raters included five military subject-matter experts (SMEs) and an ARI scientist with 28 years of research experience in military training (Table 1).

Table 1. Rater Characteristics

Rater	Rank	Key Experience
Author	Scientist	24 years, Army unit training R&D
SME1	LTC (Ret)	NTC trainee
SME2	MAJ (Ret)	NTC trainee and O/C
SME3	LTC (Ret)	NTC trainee, Vietnam veteran
SME4	MAJ (Res)	Bn FTX as S2
SME5	LTC (Res)	10 years in armor/cavalry units

2.4 Analyze the Rank Orders.

a. Spearman correlations (Rho) corrected for ties, were computed for all combinations of six raters over all BOS rankings. These indices yielded measures of agreement among the raters.[3]

[3] StatView SE was used to compute correlations. An alternative is to compute an average correlation and overall significance level, using Kendall's Concordance method. Manual computation is very complicated (Marascuilo & McSweeney, 1977). A computer program for the Concordance was not available to the author when this study was initiated. The Concordance can now can be computed on SPSS for windows.

Correlations were also computed for rankings by phase of battle. Each set of rankings was then correlated with METT-T Score to determine the relationship between performance on each BOS and mission outcome. For the validity assessment, Pearson correlations (linear and polynomial) were computed.

b. These analyses were repeated for BOS rankings by phase of battle. For example, consistency of ranking among raters for Intelligence planning was assessed. Then each rater's ranking was correlated with METT-T score to assess the relationship between intelligence planning and mission outcome. Only the five SMEs assessed the sub BOSs.

2.5 Analyze Measurement Problems

The raters reviewed each of the BOS impact statements for measurement problems. For example, they looked for characteristics of the statements that might limit reliability or relationships to mission outcome. Instructions for reviewing the statements were provided as part of Appendix D. The author did an additional in-depth analysis to extract categories of psychometric problems and to recommend solutions.

3.0 RESULTS

3.1 Analysis of Ranking Data for Overall BOS Impact Statements

a. Interrater Consistency. Interrater correlations (Spearman's Rho, corrected for ties) were computed for the three BOSs and ten sub BOSs (Appendix E). ***All*** the correlations (15 out of 15 for each overall BOS) were significant (Appendix E1). Table 2 displays mean values and ranges of values from E1.

Table 2. Mean Consistency (Rho) Of Overall BOS Rankings

BOS	Mean Rho	Range of Rhos	Range of P
Intelligence	.837	.691 to .939	.0038 to .0048
Command&Control	.939	.912 to .973	.0032 to .0062
Maneuver	.832	.682 to .969	.0036 to .0396

b. Validity

(1) In a preliminary analysis, validity coefficients for BOS rankings vs. METT-T scores were computed and scattergrams drawn for the author, SME1, and SME2. Of nine validity coefficients (3 BOSs x 3 analysts), only SME1's coefficient for Intelligence and SME2's for Maneuver proved significant. However, one exercise was clearly an outlier across the scattergrams. It was excluded from further analysis. The final analysis of all six raters showed dramatic validity results.

(2) Table 3 presents validity coefficients (linear analysis) averaged across the six raters. It also shows ranges of F-values (Fs)and alpha-error probabilities (Ps). Appendices F1 and F2 detail the individual correlations. ***All*** the validity coefficients for Maneuver and C2 are significant. ***All*** the coefficients for Intelligence are non significant.

(3) The significant validities are also relatively large. Predictive validities in personnel testing have typically accounted for about 10% to 40% of variance (Anastasi, 1982). Even the smallest coefficient(.67) in Appendices F1 and F2 accounts for 45% of variance. The largest(.96) accounts for 92% of variance.

Table 3. Mean Validity Of Rankings Vs. METT-T Score (Linear)

BOS	Mean Validity	Range of Fs	Range of Ps
Command&Control	0.76	9.44 - 15.43	.0057-.0476
Maneuver	0.82	06.94 - 77.48	.0001 - .0370

(4)Polynomial analyses are summarized in Table 4 and in Appendices F3 and F4. These validities are also consistently significant for command and control, and for maneuver with one exception out of twelve.

Table 4. Mean Validity Of Rankings Vs. METT-T Score (Polynomial)

BOS	Mean Validity	Range of Fs	Range of Ps
Command&Control	0.87	5.67 - 21.13	.0020 - .0250
Maneuver	0.87	03.74 - 34.39	.0005 - .0150

(5) In addition, the polynomial validities are higher than the linear ones, with one exception out of ten (Appendices G1 - G6). To shed some light on the import of this latter fact, Figure 1 presents a typical BOS ranking vs. METT-T score scatter plot for the Command and Control, BOS. Note, the four top-ranked exercises closely fit a linear relationship with METT-T Score. The five lowest ranked exercises show a more chaotic pattern, better suited to a curvilinear fit. These results will be examined further in the Discussion.

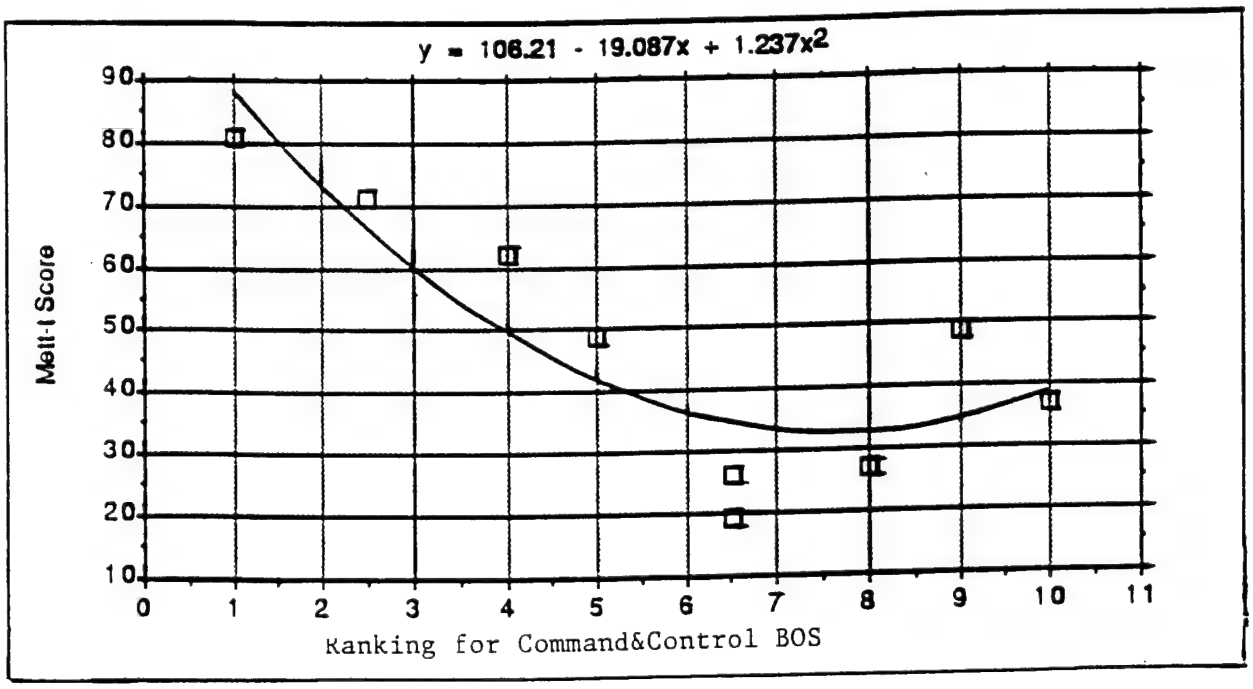

Figure 1. Representative Scatter Plot For Command And Control BOS Ranking Vs. METT-T Score.

3.2 Analysis Of BOS Impact By Phase Of Battle

a. Interrater agreement. Tables 5-7 summarize the interrater correlations (Rho, corrected for ties) for sub BOS rankings. Appendices E2 through E4 present the intercorrelation matrices.

Table 5. Summary Of Interrater Correlations For Intelligence Sub BOSs

Statistic	Planning		S2 Execution		Scout. Execution	
	Rho	Prob	Rho	Prob	Rho	Prob
Average	.774		.687		.606	
Range	.660-.884	.095-.0499	.672-.860	.0142-.0439	.415-.762	.0223-.0554
% Significant	100		70		50	

Table 6. Summary Of Interrater Correlations For Comm&Con Sub BOSs

Comparison	Planning		Preparation		Execution	
	Rho	Prob	Rho	Prob	Rho	Prob
Average	.413		.656		.834	
Range	.203-.827	.0059-.0398	.345-.892	.0243-.0352	.694-.915	.0041-.0373
% Significant	60		60		100	

Table 7. Summary Of Interrater Correlations For Maneuver Sub BOSs

Comparison	Planning		Preparation		Counter Recon		Execution	
	Rho	Prob	Rho	Prob	Rho	Prob	Rho	Prob
Average	.587		.603		.815		.758	
Range	.170-.947	.0045-0367	.354-.827	.131.-0475	.713-.951	.0043-.0323	.681-.895	.007-.0411
% Sig.	60		50		100		90	

Note the ***large variability*** of percent significant correlations for the sub BOSs, in contrast to the ***consistently*** significant intercorrelations for the overall BOSs. For example, ***100%*** of interrater correlations for maneuver/counter reconnaissance were significant, compared to ***50%*** for maneuver/preparation.[4]

b. Correlation with METT-T score. Tables 8 through 13 summarize linear and polynomial correlations between the sub-BOS rankings and METT-T scores. Details are presented in Appendices G1 - G6. Note the trends in the average correlation and in % of significant correlations. Each consistently increases from the first to the last phase of battle. The theoretical import of this observation will be addressed in the Discussion.

Table 8. Summary Of Validities Of Intelligence Sub BOS Correlations With METT-T Score: (Linear Analysis)

Comparison	Planning	S2 Execution	Sct. Execution
Average	.16	.32	.50
% Significant	0	0	20

[4] The sub BOS rankings are based on a smaller sample of unit performance - 1/3 to 1/4 of the information in the full BOS statement. Therefore, interrater reliability may be adversely influenced in accordance with the test-length principle (Anastasi, 1982, Page 114).

Table 9. Summary Of Validities Com&Control Sub BOS Correlations With METT-T Score: (Linear Analysis)

Comparison	Planning	Preparation	Execution
Average	.43	.70	.80
% Significant	0	80	100

Table 10. Summary Of Validities Of Maneuver Sub BOS Correlations With METT-T Score (Linear Analysis)

Comparison	Planning	Preparation	Counter Recon	Execution
Average	.62	.68	.61	.81
% Significant	20	80	20	100

Table 11. Summary Of Validities, Intelligence Sub BOS Correlations: (Polynomial Analysis)

Comparison	Planning	S2 Execution	Sct Execution
Average	.36	.55	.58
% Significant	0	20	20

Table 12. Summary Of Validities Of Com&Control Sub BOS Correlations With METT-T Score: (Polynomial Analysis)

Comparison	Planning	Preparation	Execution
Average	.54	.68	.85
% Significant	0	0	80

Table 13. Summary Of Validities Of Maneuver Sub BOS Correlations With METT-T Score. (Polynomial Analysis)

Comparison	Planning	Preparation	Counter Recon	Execution
Average	.72	.72	.66	.90
% Significant	20	20	20	100

c. Constraints of data quality. Along with their rankings, the SMES provided comments on the quality of the BOS statements as data elements. These comments are reproduced in Appendix H. They reveal four problems: insufficient and varying amounts of detail, unclear distinctions among sub BOSs; Ill-defined evaluative language, and inconsistent judgments within the same statements.

3.3 Scaling and Other Issues

Obstacles to Reliable, Valid, or Useful Measurement. Based on the author's review of the BOS exercise and SME comments, 6 classes of measurement problem were identified.[5]

a. Variability of dimensions, scales, and units of analysis. The BOS statements usually address the same tasks from exercise to exercise. But they often focus on different performance dimensions or use different units of analysis. One may detail problems with the intelligence preparation of the battlefield (IPB). Another may say only that more information is needed in the IPB. One may criticize spot reporting as inadequate, another may identify

[5] This section is not a criticism of O/C performance. The BOS statements are being used as a research testbed to advance knowledge about unit performance measurement for training feedback.

accuracy and timeliness as separate problems. Some statements focus only on problems and omit satisfactory instances of performance. What is excluded may be as important as what is included. Other statements judging equally good or poor exercises, include positive and negative comments.

b. "Contamination" (non independence across statements). The same performance dimensions and implicit scales may appear in statements about different BOSs. For example, ineffective repositioning of a mechanized unit was cited in the maneuver and C2 BOS statements for one particular mission.

c. Interactive effects across echelons, elements, and time. This problem can be summarized by a familiar expression. The whole is not necessarily the sum of the parts. Across phases, echelons, and elements of battle the following occurs.

(1) Good performance compensates for poor performance. Combat elements can re-position effectively, even where re-positioning was not well planned. Scouts can execute counter-reconnaissance effectively, absent planning by the S2.

(2) Good performance is offset by poor performance. Timely reporting and coordination by scouts or other teams is offset by poor follow-up by the Tactical Operations Center (TOC). Obstacles are effective, but the task force fails to mass fire.

(3) Errors by one unit creates problems for other units. Scouts who leave their proper boundaries are killed by Task Force (TF) fire because TF doesn't track adequately. The S2 analyzes the battle incorrectly because spot reports are incorrect.

(4) Finally, Performance is good or bad depending on circumstances. The TF OpOrd process takes so long that it interferes with planning by subordinates. Judged out of context, the OpOrd process might appear satisfactory.

d. Variability in types of indicators e.g., commission vs omission or 'can't do' vs 'won't do'. Both feedback and training strategy development require careful attention to alternative types of performance indicators. For example, errors of commission may require different remediation than errors of omission. A familiar example of 'can't vs won't' do is the comparison between Generals McClellan and Grant. McClellan, at the top of his West Point class 'could' but 'wouldn't. Grant at the bottom of his class had a 'dam the torpedoes' perspective.

e. Descriptive statements without evaluations. Some statements devote too much space to exercise descriptions without evaluations. Descriptions are adequately detailed in other parts of the take-home package. They detract from the impact statements, which are best used to highlight performance problems and successes.

f. Apparent contradictions. One impact statement said that a revised maneuver plan was adequate. In its next paragraph the statement cited numerous deficiencies in the plan concept. This and other contradictions are probably more apparent than real, e.g. the result of inappropriate choice of words.

4.0 DISCUSSION

4.1 Overall BOS Analysis

a. Agreement among raters

(1) The results show that BOS impact statements can be used reliably to assess relative unit performance, even though they were not designed or intended for that purpose. The data are encouraging for several reasons. First, statistical significance is unlikely for sample sizes under 30, unless population effects are extremely large (Cohen, 1992). Secondly, the ranking methods and qualifications of the analysts varied. The author, for example, used a 'check-list' approach. He assigned pluses and minuses to comments in the Impact Statements, then rank ordered on the basis of the proportion of pluses. One SME used a similar approach in conjunction with pair-wise comparisons. This variability makes the findings more generalizable. It should be possible to increase the reliability and perhaps validity even further by making the BOS measurement procedures more rigorous and uniform.

(2) One way to do so is to combine the 'checklist' approach, described above, with double pair-wise comparison. Then, experience-based judgments, not reflected in the proportion of positive statements, can be used to adjust the ranking. A complementary approach is to format and organize the BOS statement more precisely. In any case this would need to be done to derive dimensions and scales of measurements from BOS impact statements. Current research under the Joint/Multiservice Distributed Testbed (JMDT2) program suggests some ways to do this.[6]

(a) Reorganize the elements of analysis into inputs, performance processes, and outputs for each stage of battle and assess the problems for each of these components. For example, the Brigade Operations Order (OpOrd) is a critical input to Battalion Task Force (BN TF) planning. Defects here can constrain the BN TF operations. Such a break-out could facilitate the work of O/Cs specializing in BOS assessments at the National Training Center (NTC). Additional data for these reorganized elements of analysis could come from taped interviews with the players. Interviews are now sometimes conducted by O/Cs. But perhaps they can be made even more useful with the aid of a well designed form for assessing the elements of analysis discussed above.

[6] The JMDT2 program is a cooperative effort by ARI, the Naval Air Warfare Center Training Systems Division and Armstrong Laboratory to develop performance assessment methods for training feedback in Joint training.

(b) Such a form should include a scale of measurement to support diagnosis of training problems and remediations for each element of analysis of the sub BOSs. It would include quality of feedback to the unit. For example:

Intelligence: Terrain Analysis

1	2	3	4	5
Significant opportunity for improvement	Moderate opportunity for improvement	Satisfactory	Commendable	Outstanding
Comments:				

b. Relationship between BOS performance and mission outcome

(1) The results indicate that BOS performance can be related significantly and substantially to mission outcomes. Maneuver and C2 rankings predicted exercise outcome as measured by control of territory and viability of remaining blue and red forces (METT-T score components). The data failed to show that Intelligence BOS rankings predict combat outcome. This failure is counter intuitive. It also contradicts dramatic examples of effects of poor intelligence noted by Guaglianone (1992). But it is consistent with conclusions found in Crumley (1989).

(2) Crumley noted that staff planning measures, which occur early in battle, have not correlated with outcomes.[7] And, Intelligence BOS statements are heavily weighted by S2 planning. But more generally, Crumley suggested that correlations with combat outcome are increasingly difficult to establish as phase of battle is further removed from the battle's end. The sub BOS ranking data (Tables 8-10) are consistent with Crumley's suggestion. They show increasing percentages of significant correlation from planning to execution. The path from planning to METT-T outcome may be too complex for a simple correlation between the latter two phases.

(3) Crumley's view, supported by the present research, suggests the need for more sophisticated methods, designed to analyze complex, sequential events. Several methods suited to the complexities and sequential character of large unit combat are case-based expert system (CBES) technology and path analysis. CBES accommodates qualitative data in the study of complex interactions among variables. Mirabella (1993a, 1994) demonstrated such use of CBES to 'predict' combat outcomes from qualitative and quantitative variables in an historical data base. Macpherson (in preparation) has adapted path analysis to assessment of BOS

[7] Jarrett (1996) similarly failed to find a relationship between the quality of Commander's Intent and mission outcome, notwithstanding the large proportion of intent statements judged to be poorly written at NTC.

performance and its impact on combat outcomes (Macpherson & Alderman, 1994; Mirabella, 1993b).

(3) Is it possible that the validities of Maneuver and C2 result from contamination by METT-T? Do the same observations appear in the three measures? Clearly, there is contamination across these measures, but probably not enough to account for the magnitude of the validities. For example, many of the BOS statements include a comment that the Red force penetrated Blue's rear defensive boundary. This outcome would also be reflected in the METT-T score. But it would represent only one of perhaps 10 to 20 comments in any statement.

(5) What are we to make of the significance of the polynomial validity data for Maneuver and C2? Is there a true and meaningful non linearity in the data. I doubt it. The scatter plots suggest that predictions may be more random for units that do more things wrong than right in the Maneuver and C2 BOSs. The quadratic function extracts somewhat more order out this chaos than does the linear function. Combined with the scatter plots, the quadratic also illustrates that some parts of combat data are more orderly than others.

4.2 Analysis of BOS Performance by Phase of Battle

a. Interrater agreement. The results show a high degree of consistency in ranking Maneuver and C2 performance for preparation and execution. Consistency for Intelligence Planning is also high. These consistencies are noteworthy given the psychometric limitations of the impact statements. In contrast consistency is very low for the Maneuver and C2 Planning, and for Intelligence Execution. In these latter cases BOS data quality and resulting judgment problems may have been most severe. These cases may benefit most from a well designed assessment instrument.

b. Relationship to mission outcome. Unit rankings in the Preparation and Execution stages of Maneuver and C2 show consistent and strong relationships to mission outcome. Results for the Planning stages across the BOSs are inconclusive. These results are consistent with Crumly's conclusion that relationships to mission outcome are more evident for later phases of battle than for earlier phases (Crumly, 1989).

c. Data Quality

(1) The number and size of significant relationships (interrater agreements and correlations with METT-T scores) is noteworthy given BOS data quality. These suggest that with modest improvements in how the BOS statement is formatted, it could become a reliable instrument for training feedback and training strategy development. The SME raters suggested that we need more,

better organized information than is contained in the statements for consistent, accurate unit performance assessment. This recommendation is being put to a test in the JMDT2 program.

(2) How much improvement in measurement resolution is a critical question, since data collection and processing can be labor intensive. The answer depends on purposes of measurement. The modest success of the present effort suggests that even a small improvement in the format and organization of the Impact Statement could be useful to O/Cs for training feedback. For R&D and readiness reporting additional sources of information need to complement training feedback data. The outer limit would be an integration of all archive data for a rotation (e.g. dynamic plan views, detailed casualty information, video of after-action reviews). The Army-wide Standard Training After Action Review (STAARS) system may provide a vehicle for this integration.

d. Implications for Decision Support Methodology and further efforts. The results increase our confidence that BOS and phase of battle (i.e. sub BOS) may provide a useful framework and measures for use in developing a DSM and as components of a DSM. But to further develop the framework and measures we need to understand or at least explore ways to understand interactions within BOSs and relationships among BOSs including effects of synchronization. We need to move beyond analyzing static conditions (One BOS vs another) to analyzing the dynamics of exercises. Case-based reasoning methodology and path analysis are candidate tools for this purpose. Research on the use of path analysis is currently in progress at ARI (Macpherson, in preparation).

4.3 Scaling and Other Psychometric Issues

a. The results indicate opportunities for improving the BOS impact statement to serve both training feedback and decision support modeling. It generally suggests deriving from the statement computational, e.g., checklist format. The format should identify a taxonomy of performance dimensions and scales that can be applied consistently from exercise to exercise and across O/Cs. It was suggested earlier that this format should include the following: sub BOSs, elements of analysis organized into inputs, performance processes, and outputs; and a five-point rating scale to support training feedback. The need for such an instrument at the NTC has been recognized by GAO (1986).

b. A computational data base format is essential if the BOS is to be used for training feedback and for training R & D. The same can be said for video-tape and time-position replay data. In their present formats, these sources require labor intensive analysis by SMEs. Computational formats won't eliminate the need for in-depth analysis by SMEs. They would extend the value of combat training center data.

5.0 CONCLUSIONS

5.1 BOS impact statements were used reliably to evaluate the relative effectiveness of unit performance across task force exercises and phases of battle. Interrater consistency, however was less for phases of battle.

5.2 Overall rankings of the impact statements for Maneuver and C2 were shown to correlate with METT-T scores. Correlation data for Intelligence were inconclusive.

a. Relative performance across exercises for the Preparation and Execution Phases of battle for Maneuver and C2 is related to mission outcome (METT-T score).

b. Findings about the relationships between METT-T score and the Planning Phases of Maneuver, and C2 are inconclusive.

C. Findings about the relationships between METT-T and all phases of Intelligence are inconclusive.

5.3 BOS statements were inconsistent in amount of information provided and tasks evaluated.

5.4 Statements would be more useful as data base elements if they were organized formatted, and prepared more systematically and uniformly to reflect phases of battle and elements of analysis within those phases.

5.5 The BOS statements have potential for helping to develop improved methods of diagnostic feedback. The statements can also help design decision support methodology for development of training strategies.

5.6 The methods and methodology can evolve from attempts to solve the measurement problems in current BOS statements.

6.0 REFERENCES

Anastasi, A. (1982). Psychological testing. New York: Macmillan.

Cohen, J. (1992). Quantitative methods in psychology. Psychological Bulletin, 112(1), 155-159.

Crumley, L. M. (1989). Review of research and methodologies relevant to Army Command and Control performance measurement (ARI Technical Report 825). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Department of the Army. (1990). Army programs, blueprint of the battlefield (TRADOC Pam 11-9). Fort Monroe, VA: Author.

Guaglianone, D. (1992). Countersurveillance operations. military Intelligence, January-March, 15-19.

Jarrett, P. A. (1996). Analysis of combat training Center archive data for critical leader behaviors (ARI Technical Report 1038). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Kierins, J. W., Atwood, N. K., & Root, J. T. (1990). Concept for a common performance measurement system for unit training at the National Training Center (NTC) and with simulation networking (SIMNET) (ARI Research Report 1574). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A230 129)

Macpherson, D. (in preparation). Measurement of synchronization (ARI Research Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Macpherson, D., & Alderman, I. (1994, May). The role of synchronization in combat effectiveness: How do you define and measure it. Minutes and Briefings of the 8th Meeting, Department of Defense Training Technology Technical Group (T2TG).

Marascuilo, L. A., & McSweeney, M. (1977). Nonparametric and distribution-free methods for the social sciences. Monterey, CA: Brooks/Cole.

Mirabella, A. (1993a, May). Case-based expert system for combat performance analysis. Minutes and Briefings of the the Annual Meeting of the Department of Defense Training Technology Technical Group (T2TG), 292-309.

Mirabella, A. (1993b, October). The challenge of measuring group performance. In F. L. Moses (Chairman), Symposium on Issues in Providing Effective Training for Large Groups. Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting, 1248.

Mirabella, A. (1994, January). Case-based expert systems for combat performance analysis (ARI Research Product 94-03). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A285 757)

U.S. General Accounting Office. (1986). Army training, National Training Center's potential has not been realized (GAO/NSIAD-86-130). Washington, DC: Author.

7.0 APPENDICES

Appendix A. Intelligence BOS: Elements of Analysis

SUB BOS: PLANNING	SUB BOS: Execution - S2
1. Terrain analysis	1. Tracking OPFOR
2. Intel prep of the battlefield	2. Tracking TF
a. Reconnaissance plan	3. Communication with Task Force
b. Counterrecon plan	4. Battle analysis by S2
c. Counterrecon staffing[8]	5. Synchronization-integrate GSR)
d. Main battle intel plan	
3. IPB coordination: S2 and CO	SUB BOS: EXECUTION - SCOUTS
4. IPB brief	1. Positioning
a. to staff	2. Repositioning
b. to scouts (e.g. OPs)	3. Local security
5. Data collection	4. Screening - counterrecon
a. Plan	5. Tracking enemy recon units
b. Requirements/criteria	6. Hand-off of enemy recon
6. Doctrine analysis-attack form.	7. Tracking OPFOR -main battle
	a. S2
	b. Tactical Operations Ctr

[8] Wargaming by S2, S3, and S4

Appendix B. Command & Control BOS: Elements of Analysis

SUB BOS: PLANNING
1. Operations order from brigade
2. CO established concept of oper.
3. Initial planning guidance from task force commander
4. Supervision of staffing by TF CO
5. Staff coordination of combat power responsibilities
6. Staff coordination of contingencies
7. Staff coordination of combat support
8. Task Force operations order
a. timeliness
b. repositioning locations
c. repositioning criteria
d. counter-attack routes /objectives
e. coordination of ADA assets
f. fire control graphics
9. Fire support for main battle
10. Synchronization:FS, Eng., Helos
11. MOPP Management
12. Plan revision
13. Task Force opord briefing
SUB BOS: PREPARATION
1. War gaming plan
2. Repositioning rehearsal
3. TOC supervision/monitoring of defensive preparation
4. Synchronization: FS, Eng, Helos
5. Super.of survivability position
6. Refinement of planning documents
7. Counter attack rehearsal
SUB BOS: EXECUTION
1. Rearward passage of scout plt
a. Coordination by TOC of Scts&Cav
b. Coordination by Scts with Cav
c. Repositioning maneuvers

SUB BOS: EXECUTION (CONTINUED)
d. Vehicle recovery: SecArea
- offer of help by TOC
- follow through by TOC
2. Control of Screening and Counter recon battle
a. Alert units to enemy move.
b. Apply combat multiplier
c. Revise scheme of defense
d. Handoff
e. Tracking of enemy
f. Monitoring unit locations by TOC
g. Coord. of UH-60s with Intel
h. fire support
3. Counter mobility
a. analysis of hazards by TOC
b. Tracking own units by TOC
4. Control of main battle
a. Tracking of MRR by TF
b. Repositioning scts, mech tm
c. Spot reporting by teams
d. CO's view of battlefield
e. Coordination /synchron.
- Fire control
- Engineering
f. Use of company nets for C2
g. MOPP defensive activities
h. Application of combat power
I. Status keeping of Maneuver elements by TOC
- Maneuver elements
- obstacles and survivabil.

Appendix C. Maneuver BOS: Elements of Analysis

SUB BOS: PLANNING	SUB BOS: EXECUTION -MRR BATTLE
1. Maneuver planning	1. Defense vs. leading element of MRR
a. CO communicates intent	2. Tracking of MRBs
b. Defense formation,(e.g. to mass power)	3. Execution of defense in depth
c. Security formation (e.g. scout positioning)	4. Execution of deception plan
d. Use of restrictive/open terrain	5. Acquisition of MRR by TF
- to define engagement area	6. Fight battle in depth
- to mass power	7. Use of initiative by TF (e.g. preemptive strike)
e. Obstacle plan	8. View of battlefield by CO
f. Coord. of combat elements	9. Use of CAS & helos to gain initiative
g. Repositioning plans	10. Use of direct fire massing
SUB BOS: PREPARATION	11. Use of indirect fire
1. Use of available time	12. Use of obstacles and FASCAM to delay the MRR
2. Hasty defense for counter recon	13. Repositioning of combat tms
3. Control/use of bulldozers	14. Coordination (mutual support) among teams
4. Dissemination of plans (e.g.obs)	15. Status reporting by company teams
5. Wargaming	16. Synchronization of combat power
6. Rehearsal	
7. Survivability preparation	
8. Reconnaissance	
SUB BOS: EXECUTION-COUNTER RECON	
1. Defense vs. hunter-killer recon	
2. Counter recon: acquisition	
3. Counter recon: destruction	
4. Prevent enemy recon penetration	
5. Early detection/tracking of enemy	

Appendix D

INSTRUCTIONS FOR RANKING BOS IMPACT STATEMENTS

BACKGROUND

The U.S. Army Research Institute (ARI) is a field operating agency of the Deputy Chief of Staff for Personnel (DCSPER). With its headquarters in Alexandria, Virginia, it has about 200 personnel and 12 field units located at Army posts throughout the United States.

ARI has a Research and Development program to support the Army's increasing use of simulation and simulation networks to prepare units for rotations at Combat Training Centers (CTCs). A key part of the program is to develop better ways to assess training effectiveness. This assessment will help commander's decide how to get the most "bang for the buck" out of the new training technologies. It will also provide our units with the best possible training feedback.

Right now we're looking at the Take-Home Package (THP) given to the units at NTC. We're focusing initially on the impact statements for Battlefield Operating Systems. The statement is just one piece of the THP. It's prepared by the O/Cs to summarize lessons learned within each BOS. Its purpose is to help the unit plan follow-up training at home station. The question we're asking is how can we use it to characterize unit performance for research and development to improve training strategies. We need your help in analyzing a sample of statements from 10 exercises. We think the answers can help the Army make better use of very promising technology.

INSTRUCTIONS

1. The data sets provided include BOS statements for 10 exercises, played by 10 different battalion task force units at NTC. For each exercise there is an impact statement for Intelligence, Command and Control, and Maneuver. In the top right corner of each statement is an identifying number (1 to 10). The number has no significance except to identify the exercise. We've re-ordered the I.D.s from one set to the next to prevent a comparison across BOS statements. At this point, we're looking for independent judgments of each statement. At later stages of the research, we want to look at systematic ways to combine different pieces of information to arrive at diagnostic judgments.

2. You'll find that the statements cut across many dimensions, including echelon, unit element, and phase of battle (i.e., planning, preparation, execution). They may contain both positive and diagnostic evaluations. For this initial phase of research, we're interested in overall, relative judgments. On balance is the BOS performance and impact better or worse for one exercise than another?

3. Taking one set at a time, e.g. command and control, please read all ten statements. Then rank the statements from most effective performance (#1) to least effective performance (#10). Use a method called Double Pair-Wise Comparison. Here's how:

a. Compare Statement 1 with each of the other 9 statements. Decide which statement of each pair shows better overall performance. Indicate that statement number in Row 1, Columns 2-10 of the table for that BOS. If the statements are tied, write "T".

b. Next compare Statement 2 with each of the other 9 statements (including Statement 1). Again, decide which of each pair shows better overall performance. Indicate that statement number in Row 2, Column 1 and Columns 3 - 10.

c. Repeat for Statements (i.e. Rows) 3 through 10.

d. Review and revise your judgments as often as you want, especially where statements are tied.

e. A filled-in ranking form (Example 1), at the end of these instructions, illustrates a table filled in with pair-wise ranking data. These are made-up data.

4. Repeat the procedures for sub BOSs (phases of battle within BOSs. For example, after you finish the overall ranking for the Intelligence BOS, rank the statements for intelligence-planning, intelligence-execution (S2), and intelligence-execution (scouts).

Use the pre-labeled forms which we have provided to document your comparisons. The BOSs and their phases of battle are:

a. Intelligence: Planning, Execution (S2), Execution (Scouts)

b. Command & Control: Planning, Preparation, Execution

c. Maneuver: Planning, Preparation, Counter Reconnaissance, Execution.

5. To help you judge the exercise statements, we've compiled lists of elements of analysis for each BOS. [Lists were included in original instruction package. In this report, they appear in Appendices A, B, and C]. The lists were derived by retired military officers from the BOS Impact Statements.

6. We welcome your reactions and comments, especially on the analysis methods and on the impact statements. Please use the blank sheet provided to write any comments you may have. We welcome your reactions.

7. How to compute final ranks from the completed table.

a. Count the number of times each statement is judged superior. Count down the column representing each exercise and across its row. Don't count the "Ts".

> Example: Turn to Example 1 (at the end of these instructions). Going across Row 1, count the number of times Statement 1 appears. The count = 0. Enter it as the total for Row 1. Now, going down Column 1, count the number of times Statement 1 is judged superior. Again the count = 0. Enter it as the total for Column 1. Add the Row 1 and Column 1 totals to get a net value for Statement 1. In this case the net value = 0. The row and column totals may not be the same.

b. In the appropriate table for each BOS or sub BOS (Look at Example 2) list the statement numbers and net values from highest to lowest. If two statements have the same net value, examine how each exercise did during the two comparisons of each to the other, if one of them "bettered" the other twice, or once and tied the other time, it is ranked ahead. Otherwise (when each battle ranks "better" than the other just once) list it as a tie. Average the two positions in question and assign each exercise that average value.

Example: Look at the illustration in Example 2. Exercises 8 and 10 are assigned the rank of 6.5. Their net values were both 8. But in one comparison 8 was judged superior to 10. In the second comparison, 10 was judged superior to 8. Therefore these are listed as ties. They jointly occupy Ranks 6 and 7. So, give each a 6.5 rank.

c. Comments on how you did the ratings or reactions to the impact statements would be very helpful. Please make any notes you care to on the ranking forms. For example, were there parts of the statements you weighted more heavily than other parts? What difficulties, if any, did you have in making judgments?

8. For data analysis purposes please indicate the following:

a. Social Security Number (Last four digits) _____

b. Your Rank/Branch_____

c. Combat Training Center (CTC) Experience (e.g., trainee, observer/controller, number of rotations.)

National Training Center (NTC)_____
_____
_____
_____

Joint Training Readiness Center (JTRC)_____
_____
_____
_____

Combat Maneuver Training Center (CMTC)_____
_____
_____
_____

Battle Command Training Program (BCTP)_____
_____
_____
_____

d. Combat Experience (e.g. Desert Shield/Storm, Panama, Granada).
_____
_____
_____
_____

Example 1. Illustration of completed ranking form. SSN (last 4 digits): __ __ __ __

Statement ID and Row Number	Statement I.D. & Column Numbers <BOS or Sub BOS Title>										Row Totals (R)
	1	2	3	4	5	6	7	8	9	10	
1		2	3	4	5	6	T	8	9	10	0
2	2		3	2	2	6	7	2	2	10	5
3	3	2		4	3	6	3	8	3	3	5
4	4	4	3		5	6	4	4	4	4	6
5	5	5	5	5		6	4	5	5	5	7
6	6	1	6	6	6		6	6	6	6	8
7	7	7	3	4	5	6		8	9	10	2
8	8	2	8	4	5	6	8		8	8	5
9	9	2	3	4	5	6	9	9		9	4
10	10	10	3	4	5	6	10	10	9		4
Column (C) Totals	0	4	6	6	6	9	1	3	3	4	
Net Value (R + C)											

Example 2

Rank order before adjustment	1	2	3	4	5	6	7	8	9	10
Exercise I.D.	6	5	4	3	2	10	8	9	7	1
Net Value	17	13	12	11	9	8	8	6	3	0
Rank order with adjustment	1	2	3	4	5	6.5	6.5	8	9	10

[Pre-labeled blank forms were included in original instruction package for each BOS and sub BOS.]

Appendix E. Interrater Correlations for BOSs and Sub BOSs

Appendix E1. Intercorrelation Table: Spearman Rho Corrected For Ties

Comparison	Intelligence		Com&Control		Maneuver	
	Rho	Prob	Rho	Prob	Rho	Prob
Author vs E1	.897	.0071	.915	.0060	.744	.0256
E2	.803	.0160	.939	.0048	.766	.0216
E3	.812	.0153	.945	.0046	.686	.0396
E4	.932	.0052	.973	.0035	.884	.008
E5	.882	.0081	.930	.0053	.765	.0218
E1 vs E2	.827	.0131	.921	.0057	.802	.0161
E3	.939	.0048	.927	.0054	.793	.0174
E4	.881	.0082	.942	.0047	.945	.0046
E5	.875	.0087	.955	.0042	.969	.0036
E2 vs E3	.845	.0112	.982	.0032	.912	.0062
E4	.691	.0382	.912	.0062	.912	.0062
E5	.861	.0098	.924	.0056	.788	.0181
E3 vs E4	.739	.0265	.927	.0054	.875	.0087
E5	.728	.0290	.945	.0046	.713	.0325
E4 vs E5	.856	.0102	.955	.0042	.927	.0054
Sum	12.568	0.2114	14.092	0.0734	12.481	0.2354
Average	.837		.939		.832	
Range	.691 - .939	.038 - .0048	.912 - .973	.0062 - .0032	.686 - .969	.0036 - .0396

Appendix E2. Intercorrelation Table: Intelligence Sub BOS Rankings (Spearman Corrected For Ties)

Comparison	Planning		S2 Execution		Scout Execution	
	Rho	Prob	Rho	Prob	Rho	Prob
E1 vs E2	.666	.0458	.672	.0439	.694	.0374
E3	.777	.0214	.774	.0203	.639	.0554
E4	.864	.0095	.577	NS	.500	NS
E5	.809	.0152	.715	.0320	.751	.0243
E2 vs E3	.832	.0126	.815	.0144	.762	.0223
E4	.778	.0195	.860	.0099	.663	.0436
E5	.660	.0479	.667	.0455	.415	NS
E3 vs E4	.884	.0080	.817	.0142	.449	NS
E5	.723	.0300	.505	NS	.623	NS
E4 vs E5	.743	.0258	.463	NS	.564	NS
Sum	7.736		6.865		6.06	
Average	.774		.687		.606	
Range	.660-.884	.095-.0499	.672-.860	.0142-.0439	.415-.762	.0223-.0554

Appendix E3. Intercorrelation Table: Comm&Con Sub BOS Rankings (Spearman Corrected For Ties)

Comparison	Planning		Preparation		Execution	
	Rho	Prob	Rho	Prob	Rho	Prob
E1 vs E2	.827	.0131	.751	.0243	.915	.0060
E3	.917	.0059	.702	.0352	.899	.0070
E4	-.128	NS	.637	NS	.850	.0108
E5	.773	.0204	.516	NS	.780	.0194
E2 vs E3	.827	.0131	.732	.0280	.873	.0088
E4	-.203	NS	.709	.0333	.877	.0085
E5	.685	.0398	.707	.0340	.776	.0213
E3 vs E4	-.312	NS	.892	.0074	.957	.0041
E5	.709	.0335	.345	NS	.694	.0373
E4 vs E5	.031	NS	.568	NS	.715	.0320
Sum	4.126		6.559		8.336	
Average	.413		.656		.834	
Range	.-203 - .827	.0059.- 0398	.345- .892	.0243- .0352	.694- .915	.0041- .0373

Appendix E4. Intercorrelation Table: Maneuver Sub BOS Rankings (Spearman Corrected For Ties)

Comparison	Planning		Preparation		Counter Reconnaissance		Execution	
	Rho	Prob	Rho	Prob	Rho	Prob	Rho	Prob
E1 vs E2	.947	.0045	.737	.0271	.835	.0123	.895	.0072
E3	.762	.0222	.354	NS	.713	.0323	.708	.0337
E4	.696	.0367	.360	NS	.724	.0297	.737	.0271
E5	.393	NS	.506	NS	.951	.0043	.591	NS
E2 vs E3	.843	.0114	.566	NS	.888	.0078	.681	.0411
E4	.737	.0271	.724	.0299	.762	.0222	.865	.0094
E5	.170	NS	.745	.0255	.872	.0089	.737	.027
E3 vs E4	.762	.0222	.661	.0475	.911	.0063	.738	.027
E5	.311	NS	.552	NS	.778	.0196	.800	.0163
E4 vs E5	.245	NS	.827	.0131	.719	.031	.830	.0127
Sum	5.866		6.032		8.153		7.582	
Average	.587		.603		.815		.758	
Range	.170- .947	.0045- 0367	.354- .827	.131.- 0475	.713- .951	.0043- .0323	.681- .895	.0072- .0411

Appendix F. Correlations of BOS Rankings and METT-T Scores

Appendix F1. Validity Of Comm&Control Rankings (Linear Analysis)

BOS	Validity	F-value	P <
Author	.81	13.50	.0080
SME1	.72	07.71	.0274
SME2	.76	09.44	.0200
SME3	.77	10.46	.0144
SME4	.83	15.43	.0057
SME5	.67	05.75	.0476
Average	0.76	10.38	0.0205
Range	.67 - .81	9.44 - 15.43	.0057-.0476

Appendix F2. Validity Of Maneuver Rankings (Linear Analysis).

BOS	Validity	F-value	P <
Author	.71	06.94	.0370
SME1	.79	11.86	.0108
SME2	.89	10.57	.0014
SME3	.96	77.48	.0001
SME4	.84	16.72	.0051
SME5	.73	07.81	.0268
Average	0.82	21.9	0.0135
Range	.73 - .96	06.94 - 77.48	.0001 - .0370

Appendix F3. Validity Of Comm&Control Rankings(Polynomial Analysis).

BOS	Validity	F-value	P <
Author	.85	07.64	.0224
SME1	.85	07.64	.0222
SME2	.94	21.13	.0020
SME3	.90	12.82	.0068
SME4	.86	8.85	.0162
SME5	.81	05.67	.0250
Average	0.87	10.63	0.0158
Range	.81 - .90	5.67 - 21.13	.0020 - .0250

Appendix F4. Validity Of Maneuver Rankings (Polynomial Analysis).

BOS	Validity	F-value	P <
Author	.74	03.74	NS
SME1	.87	08.95	.0150
SME2	.89	11.58	.0090
SME3	.96	34.39	.0005
SME4	.88	10.25	.0116
SME5	.87	09.44	.0140
Average	0.87	13.06	0.01
Range	.74 - .96	03.74 - 34.39	.0005 - .0150

Appendix G. Correlations of Sub BOS Rankings with METT-T Scores

Appendix G1. Validities For Intelligence Sub BOS Rankings: (Linear Analysis)

Comparison	Planning		S2 Execution		Scout Execution	
	Rho	Prob	Rho	Prob	Rho	Prob
Mett-t vs E1	.28	NS	.25	NS	.32	NS
E2	.18	NS	.27	NS	.81	.0081
E3	.12	NS	.25	NS	.60	NS
E4	.07	NS	.45	NS	.67	.500
E5	.14	NS	.39	NS	.09	NS
Average	.16		.32		.50	
Range	.07-.28		.25-.45		.09-.81	

Appendix G2. Validities For Com&Control Sub BOS Rankings: (Linear Analysis)

Comparison	Planning		Preparation		Execution	
	Rho	Prob	Rho	Prob	Rho	Prob
Mett-t vs E1	.57	NS	.73	.0249	.80	.0091
E2	.52	NS	.71	.0316	.85	.0035
E3	.61	NS	.76	.018	.85	.0039
E4	.07	NS	.59	NS	.77	.0150
E5	.36	NS	.76	.0175	.73	.0245
Average	.43		.71		.8	
Range	.07-.61		.59-.76		.73-.85	

Appendix G3. Validities For Maneuver Sub BOS Rankings (Linear Analysis)

Comparison	Planning		Preparation		Counter Recon		Execution	
	Rho	Prob	Rho	Prob	Rho	Prob	Rho	Prob
Mett-t vs E1	.59	NS	.32	NS	.44	NS	.70	.0357
E2	.60	NS	.68	.0431	.58	NS	.78	.0131
E3	.70	.0362	.75	.0198	.75	.0198	.75	.0198
E4	.84	.0049	.96	.0001	.76	.0176	.94	.0002
E5	.39	NS	.77	.0162	.52	NS	.88	.0016
Sum	3.12		3.48		3.05		4.05	
Average	.62		.70		.61		.81	
Range	.59-.84		.32-.96		.44-.76		.70-.94	

Appendix G4. Validities For Intelligence Sub BOS Rankings: (Polynomial Analysis)

Comparison	Planning		S2 Execution		Scout Execution	
	Rho	Prob	Rho	Prob	Rho	Prob
Mett-t vs E1	.460	NS	.43	NS	.65	NS
E2	.360	NS	.46	NS	.83	.0305
E3	.560	NS	.52	NS	.60	NS
E4	.070	NS	.47	NS	.68	NS
E5	.360	NS	.87	.0151	.15	NS
	1.81		2.75		2.91	
Average	.36		.55		.58	
Range	.07-.56		.43-.87		.15-.83	

Appendix G5. Validities For Com&Control Sub BOS Rankings: (Polynomial Analysis)

Comparison	Planning		Preparation		Execution	
	Rho	Prob	Rho	Prob	Rho	Prob
Mett-t vs E1	.61	NS	.75	NS	.89	.0101
E2	.59	NS	.71	NS	.91	.0054
E3	.72	NS	.78	NS	.87	.0159
E4	.38	NS	.75	NS	.83	.0306
E5	.41	NS	.39	NS	.73	NS
Average	.54		.68		.85	
Range	.38-.72		.39-.78		.73-.91	

Appendix G6. Validities For Maneuver Sub BOS Rankings (Polynomial Analysis)

Comparison	Planning		Preparation		Counter Recon		Execution	
	Rho	Prob	Rho	Prob	Rho	Prob	Rho	Prob
Mett-t vs E1	.61	NS	.32	NS	.470	NS	.93	.0028
E2	.62	NS	.69	NS	.590	NS	.88	.0112
E3	.84	.026	.85	.0215	.850	.0215	.85	.0215
E4	.84	.027	.97	.0002	.840	.0265	.94	.0012
SE5	.67	NS	.78	NS	.550	NS	.90	.0061
Sum	3.58		3.61		3.30		4.503	
Average	.72		.72		.67		.90	
Range	.61-.84		.32-.97		.47-..85		.85-.94	

Appendix H. Verbatim Report of Raters on Quality of BOS Statements

1. Too often there was too little information available and one is faced with "inferring" from other observations the quality of performance in one of the sub BOSs.

2. It is not always easy to separate planning from execution. The latter is very dependent on the former and when trying to infer how well planning was done sometimes we had to use what was reported regarding execution.

3. The choice of O/C words are subject to interpretation. One O/C may mean that performance was acceptable but not perfect when he says needs improvement", while another may mean the unit is completely deficient.

4. In the area of planning, it is not clear if we should be evaluating the "quality of the plan" or the "performance of the planning process". These could be very different.

5. When a unit has improved it is a matter of interpretation as to their basic performance relative to another unit.

6. Some write-ups are internally inconsistent, stating that overall preparation was adequate but then listing some preparation areas (e.g., engineers) that were not adequately performed.

7. When performance is unsatisfactory in a certain sub BOS it is difficult to pair-wise compare units. In areas where there seem to be wide spread problems (e.g., Command and Control Preparation) score distribution appears bimodal.

8. Last, we think you have taken this analysis based on O/C BOS reports to as fine a level of detail as is feasible (perhaps too fine in some cases). We would not recommend that you attempt to further assess these below the sub BOSs used here using only O/C written reports.